

**Taller #2  
Muestreo  
Econometría 06169**

**Profesor: Julio César Alonso**  
**Monitores: Paúl Semaán**  
**Francisco Quevedo**

**Notas:**

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 30 de enero del 2006.

**INSTRUCCIONES:**

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

1. Teniendo en cuenta la información del archivo T2-01-06, que corresponde a una población de 50000 personas adultas de una pequeña isla mediterránea y que contiene su ingreso mensual (medido en salarios mínimos) y su género (M: masculino, femenino: F) resuelva:
  - a. Teniendo en cuenta que la variable de interés es el ingreso, suponga que los primeros 30 datos corresponden a la muestra piloto, que son elementos seleccionados intencionalmente teniendo en cuenta los criterios de expertos estadísticos. Partiendo de esta muestra determine el tamaño mínimo de una muestra aleatoria simple para estimar el ingreso total de la comunidad con un error absoluto no mayor de 10000 salarios mínimos y una confiabilidad mínima del 95%.
  - b. Siendo la variable de interés el ingreso, ahora genere usted una muestra aleatoria de tamaño 20 (muestra piloto 2) y a partir de ésta halle el tamaño de muestra óptimo, tenga en cuenta que el tamaño de la muestra que usted va a generar es inferior al que consideran los expertos estadísticos de la isla como proporcional al tamaño de la población.
  - c. Con base en la muestra piloto sugerida por los expertos, determine el tamaño de la muestra con el fin de determinar la proporción de hombres con un error absoluto no mayor de 0.04 y una confiabilidad del 90%.
2. Teniendo en cuenta que la variable de interés es el ingreso y empleando la misma información:
  - a. Halle el tamaño de la muestra con la misma precisión absoluta y confiabilidad de los puntos anteriores asumiendo que lo único que se conoce es que ninguna persona de la población tiene un ingreso superior a 16 salarios mínimos y que existen personas que al encontrarse desempleadas no devengan ningún salario mínimo y que a su vez la distribución de ingresos de la población es aproximadamente normal.
  - b. Halle el tamaño de la muestra con la misma precisión absoluta y confiabilidad de los puntos anteriores asumiendo que lo único que se

conoce es que ninguna persona de la población tiene un ingreso superior a 16 salarios mínimos y que existen personas que al encontrarse desempleadas no devengan ningún salario mínimo y es posible decir que la distribución que tienen los ingresos es uniforme. Compare el tamaño de muestra con el obtenido en el literal a. ¿A que se puede deber su semejanza o diferencia?

3. Con base en los datos del archivo T2-01-06 responda:
  - a. Teniendo en cuenta que la variable de interés es el ingreso y que las clases socioeconómicas de la isla se dividen en alta (individuos con ingresos mayores a diez salarios mínimos), clase media (individuos con ingresos mayores a siete salarios mínimos, pero inferiores a 10), clase baja (individuos con menos de siete salarios mínimos). Genere una muestra de veinte individuos para cada una de las clases socioeconómicas y a partir de éstas determine el tamaño mínimo de la muestra global para estimar el ingreso total de la comunidad con un error absoluto no mayor de 13000 salarios mínimos y una confiabilidad mínima del 95%. Además a partir de éstas tres muestras estime la media **y la varianza** poblacional
  - b. Determine el tamaño de la muestra del estrato "alto" "medio" y "bajo" para estimar el ingreso total de la comunidad con un error absoluto no mayor de 13000 salarios mínimos y una confiabilidad mínima del 95
4. Continuando con la pregunta anterior:
  - a. Con base en los ingresos de total de la población, elabore una muestra aleatoria simple del tamaño óptimo hallado en el literal a. del punto 1, estime la media y la desviación típica y luego determine un intervalo de confianza del 95% para esta media. Repita este procedimiento diez veces y elabore una tabla donde muestre sus resultados (intervalos, medias y desviaciones típicas para cada una de las 10 muestras aleatorias). Calcule **y compare** el valor real de la media poblacional (**calculado con el total de la población**) con las medias estimadas en este literal y utilice este valor para analizar los resultados obtenidos en este literal.
  - b. Con base en los ingresos del total de la población, elabore una muestra aleatoria estratificada del tamaño global óptimo hallado en el literal a. del punto 3 y a su vez con el tamaño por estrato hallado en el literal b. del mismo punto, estime la media y la desviación típica y luego determine un intervalo de confianza del 95% para esta media. Repita este procedimiento diez veces y elabore una tabla donde muestre sus resultados (intervalos, medias y desviaciones típicas). Calcule **y compare** el valor real de la media poblacional (**calculado con el total de la población**) con las medias estimadas en este literal y utilice este valor para analizar los resultados obtenidos en este literal.
5. Teniendo en cuenta la distribución de los ingresos poblacionales, y basándose en este explique los resultados obtenidos en el punto cuatro. Justifique suficientemente su respuesta.

**Taller #2  
Respuestas Sugeridas  
Muestreo  
Econometría 06169**

**Profesor: Julio César Alonso**  
**Monitores: Paúl Semaán**  
**Francisco Quevedo**

**Notas:**

- o Recuerde que sólo dos preguntas, seleccionadas al azar, serán calificadas.
- o Este taller es para ser entregado en los primeros 10 minutos de la clase del próximo 30 de enero del 2006.

**INSTRUCCIONES:**

- Este taller debe ser escrito en computador y entregado en papel.
- Cuando sea posible, debe mostrar el procedimiento efectuado para llegar a sus resultados.

1. Teniendo en cuenta la información del archivo T2-01-06, que corresponde a una población de 50000 personas adultas de una pequeña isla mediterránea y que contiene su ingreso mensual (medido en salarios mínimos) y su género (M: masculino, femenino: F) resuelva:

- a. Teniendo en cuenta que la variable de interés es el ingreso, suponga que los primeros 30 datos corresponden a la muestra piloto, que son elementos seleccionados intencionalmente teniendo en cuenta los criterios de expertos estadísticos. Partiendo de esta muestra determine el tamaño mínimo de una muestra aleatoria simple para estimar el ingreso total de la comunidad con un error absoluto no mayor de 10000 salarios mínimos y una confiabilidad mínima del 95%.

Una forma de estimar la suma total de los ingresos de la población es por medio de la expresión  $n \cdot \mu$ . Así, dado que conocemos el tamaño de la población, tendremos que estimar la media poblacional ( $\mu$ ). Así, todos nuestros cálculos deben expresarse en términos de la estimación de la media. Noten que en este caso tenemos que:

$$z = 1.96$$

$$\delta = 0.2 = (10000/50000)$$

$$S^2 = 7.76281392$$

$$N = 50000$$

(La varianza se calcula a partir de la muestra piloto).

Así, el tamaño de la MAS sin reposición necesario para garantizar las condiciones requeridas es:

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left( \frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{\frac{1.96^2 \times 7.76281392}{0.2^2}}{1 + \frac{1}{50000} \left( \frac{1.96^2 \times 7.76281392}{0.2^2} \right)} = 734.587355 \approx 735$$

Es decir, el tamaño de muestral debe ser de 735 y no se deben incluir los individuos que pertenecieron a la prueba piloto.

- b. Siendo la variable de interés el ingreso, ahora genere usted una muestra aleatoria de tamaño 20 (muestra piloto 2) y a partir de ésta halle el tamaño de muestra óptimo, tenga en cuenta que el tamaño de la muestra que usted va a generar es inferior al que consideran los expertos estadísticos de la isla como proporcional al tamaño de la población.

En este caso: Noten que los siguientes resultados pueden diferir un poco pues los datos han sido generados aleatoriamente

$$z = 1.96$$

$$\delta = 0.2 = (10000/50000)$$

$$S^2 = 9.32448031$$

$$N = 50000$$

Por tanto, el tamaño de la MAS sin reposición necesaria para garantizar las condiciones requeridas es:

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left( \frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{\frac{1.96^2 \times 9.32448}{0.2^2}}{1 + \frac{1}{50000} \left( \frac{1.96^2 \times 9.32448}{0.2^2} \right)} = 879.76 \approx 880$$

Es decir, el tamaño de la muestra debe ser de 880.

- c. Con base en la muestra piloto sugerida por los expertos, determine el tamaño de la muestra con el fin de determinar la proporción de hombres con un error absoluto no mayor de 0.04 y una confiabilidad del 90%.

Con base en la prueba piloto de 30 datos es posible determinar que  $\hat{p} = \frac{15}{30} = 0.5$  con lo cual es posible determinar el tamaño de la muestra:

$$n = \frac{\frac{z^2 \hat{p} \hat{q}}{\delta^2}}{\frac{N-1}{N} + \frac{1}{N} \times \frac{z^2 \hat{p} \hat{q}}{\delta^2}} = \frac{\frac{1.645^2 \times 0.5 \times 0.5}{0.04^2}}{\frac{49999}{50000} + \frac{1}{50000} \times \frac{1.645^2 \times 0.5 \times 0.5}{0.04^2}} = 419.279 \approx 419$$

Es decir, el tamaño de la muestra debe ser de 419.

2. Teniendo en cuenta que la variable de interés es el ingreso y empleando la misma información:

- a. Halle el tamaño de la muestra con la misma precisión absoluta y confiabilidad de los puntos anteriores asumiendo que lo único que se conoce es que ninguna persona de la población tiene un ingreso superior a 16 salarios mínimos y que existen personas que al encontrarse desempleadas no devengan ningún salario mínimo y que a su vez la distribución de ingresos de la población es aproximadamente normal.

Para responder a este punto es necesario tener en cuenta algunas características de la distribución normal, la más importante es que entre la media y mas o menos tres desviaciones estándar se encuentran el 99.73% de los valores, lo cual por simplicidad lo tomamos como el 100%. Luego se tiene que el rango de la variable ingreso es igual a la diferencia entre los valores extremos es decir de 16 salarios mínimos y al igualar este valor al valor del rango de 6 desviaciones estándar se puede encontrar la varianza muestral:

$6S = 16$      $S = 2.6666$  y por tanto  $S^2 = 7.1111$  con lo cual se tiene el siguiente tamaño de muestra:

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left( \frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{\frac{1.96^2 \times 7.1111}{0.2^2}}{1 + \frac{1}{50000} \left( \frac{1.96^2 \times 7.1111}{0.2^2} \right)} = 673.74 \approx 674$$

Es decir, el tamaño de la muestra debe ser de 674.

- b. Halle el tamaño de la muestra con la misma precisión absoluta y confiabilidad de los puntos anteriores asumiendo que lo único que se conoce es que ninguna persona de la población tiene un ingreso superior a 16 salarios mínimos y que existen personas que al encontrarse desempleadas no devengan ningún salario mínimo y es posible decir que la distribución que tienen los ingresos es uniforme. Compare el tamaño de muestra con el obtenido en el literal a. ¿A que se puede deber su semejanza o diferencia?

Para responder a este punto es necesario utilizar las propiedades de la distribución uniforme para estimar la varianza. Al ser 16 salarios mínimos el mayor valor que toma la variable ingreso y el menor cero entonces la varianza de esta distribución está dada por:

$$\frac{(b-a)^2}{12} \text{ reemplazando se obtiene que: } S^2 = \frac{(16-0)^2}{12} = 21.33 \text{ por lo tanto:}$$

$$n = \frac{\frac{z_{\alpha/2}^2 S^2}{\delta^2}}{1 + \frac{1}{N} \left( \frac{z_{\alpha/2}^2 S^2}{\delta^2} \right)} = \frac{\frac{1.96^2 \times 21.33}{0.2^2}}{1 + \frac{1}{50000} \left( \frac{1.96^2 \times 21.33}{0.2^2} \right)} = 1968.20 \approx 1968$$

Es decir, el tamaño de la muestra debe ser de 1968, el incremento en el tamaño de la muestra se puede deber a la heterogeneidad de la población inferida.

3. Con base en los datos del archivo T2-01-06 responda:

- a. Teniendo en cuenta que la variable de interés es el ingreso y que las clases socioeconómicas de la isla se dividen en alta (individuos con ingresos mayores a diez salarios mínimos), clase media (individuos con ingresos mayores a siete salarios mínimos, pero inferiores a 10), clase baja (individuos con menos de siete salarios mínimos). Genere una muestra de veinte individuos para cada una de las clases socioeconómicas y a partir de éstas determine el tamaño mínimo de la muestra global para estimar el ingreso total de la comunidad con un error absoluto no mayor de 13000 salarios mínimos y una confiabilidad mínima del 95%. Además a partir de éstas tres muestras estime la media y la varianza poblacional

Para el cálculo del tamaño de la muestra global se tiene que:

$$\begin{aligned} z &= 1.96 \\ \delta &= 0.26 = (13000 / 50000) \\ S_{rico} &= 0.343 \\ S_{medio} &= 0.91 \\ S_{pobre} &= 1.71 \\ N &= 50000 \end{aligned}$$

Reemplazando se tiene que:

$$n = \frac{\frac{z_{\alpha/2}^2 \left( \sum_{h=1}^H W_h S_h \right)^2}{\delta^2}}{1 + \frac{1}{N} \left( \frac{z_{\alpha/2}^2 \left( \sum_{h=1}^H W_h S_h \right)^2}{\delta^2} \right)} = \frac{\frac{1.96^2 \times (0.6 \times 1.71 + 0.29 \times 0.91 + 0.10 \times 0.34)^2}{0.26^2}}{1 + \frac{1}{50000} \left( \frac{1.96^2 \times (0.6 \times 1.71 + 0.29 \times 0.91 + 0.10 \times 0.34)^2}{0.26^2} \right)} = 99.40 \approx 99$$

Es decir el tamaño de la muestra global es de 99 individuos.

Para estimar la media poblacional con los datos del punto 3a. se tiene en cuenta las ponderaciones y las medias de cada estrato:

$$\bar{y} = \sum_{h=1}^H W_h \bar{y}_h = 0.1 \times 10.51 + 0.29 \times 8.38 + 0.6 \times 3.95 = 5.95, \text{ posteriormente se estima la}$$

varianza como se muestra a continuación:

$$Var(\bar{y}) = \sum_{h=1}^H W_h^2 Var(\bar{y}_h) = 0.1^2 \times 0.00586 + 0.29^2 \times 0.0414 + 0.6^2 \times 0.1473 = 0.057$$

- b. Determine el tamaño de la muestra del estrato “alto” “medio” y “bajo” para estimar el ingreso total de la comunidad con un error absoluto no mayor de 13000 salarios mínimos y una confiabilidad mínima del 95

A partir de los datos obtenidos en el literal a. es posible determinar el tamaño de cada estrato:

$$n_{rico} = n \frac{W_h S_H}{\sum_{h=1}^H W_h S_H} = 99 \frac{0.34 \times 0.1}{0.6 \times 1.71 + 0.29 \times 0.91 + 0.10 \times 0.34} = 2.54 \approx 3$$

$$n_{medio} = n \frac{W_h S_H}{\sum_{h=1}^H W_h S_H} = 99 \frac{0.29 \times 0.91}{0.6 \times 1.71 + 0.29 \times 0.91 + 0.10 \times 0.34} = 19.73 \approx 20$$

$$n_{pobre} = n \frac{W_h S_H}{\sum_{h=1}^H W_h S_H} = 99 \frac{0.6 \times 1.71}{0.6 \times 1.71 + 0.29 \times 0.91 + 0.10 \times 0.34} = 76.71 \approx 77$$

Es decir, los tamaños de las muestras para los estratos “rico”, “medio” y “pobre” deben ser de 3, 20, 77 individuos respectivamente.

4.

- a. Con base en los ingresos de toda la población, elabore una muestra aleatoria simple del tamaño óptimo hallado en el literal a. del punto 1, estime la media y la desviación típica y luego determine un intervalo de confianza del 95% para esta media. Repita este procedimiento diez veces y elabore una tabla donde muestre sus resultados (intervalos, medias y desviaciones típicas para cada una de las 10 muestras aleatorias). Calcule el valor real de la media poblacional obtenido con el total de la población con la media estimada y utilice este valor para analizar los resultados obtenidos en este literal.

Como se aprecia en la tabla 1. la media estimada es una muy buena aproximación de la media real de la población que es de 6.006, además en nueve de las diez estimaciones, la media poblacional se encuentra dentro del intervalo correspondiente lo cual refleja que el muestreo aleatorio simple es un buen método en este caso, ya que el valor estimado es muy cercano al valor real. Otro aspecto importante es que el intervalo es bastante pequeño, lo cual denota un gran conocimiento de la población a través de la muestra.

Tabla 1. Resultados de las diez pruebas utilizando MAS

Réplica	Media	Error Estandar	Intervalo	
			Lim inferior	Lim superior
1	6.00161291	2.885876635	5.762531461	6.240694366
2	6.04073112	2.972152267	5.794502137	6.286960111
3	6.04203188	2.892655544	5.802388823	6.28167493
4	5.99867247	2.892783591	5.759018804	6.238326127
5	5.74920989	2.882926933	5.510372809	5.988046975
6	6.19519008	2.940799415	5.951558526	6.438821624
7	6.12083274	2.8885014	5.881533843	6.360131646
8	5.9403795	2.816085162	5.707079943	6.17367905
9	6.1651452	2.919674527	5.923263755	6.407026655
10	6.04628952	2.959448196	5.801113006	6.291466034

- b. Con base en los ingresos del total de la población, elabore una muestra aleatoria estratificada del tamaño global óptimo hallado en el literal a. del punto 3 y a su vez con el tamaño por estrato hallado en el literal b. del mismo punto, estime la media y la desviación típica y luego determine un intervalo de confianza del 95% para esta media. Repita este procedimiento diez veces y elabore una tabla donde muestre sus resultados (intervalos, medias y desviaciones típicas). Calcule el valor real de la media poblacional obtenido con el total de la con la media estimada y utilice este valor para analizar los resultados obtenidos en este literal.

Como se aprecia en la tabla 2. la media estimada es una muy mala aproximación de la media real de la población que es de 6.006, además en siete de las diez estimaciones, la media poblacional se encuentra fuera del intervalo correspondiente lo cual refleja que el muestreo aleatorio estratificado en este caso no es el método idóneo, ya que el valor estimado no se aproxima al valor real.

Tabla 2. Resultados de las diez pruebas utilizando MEA

Réplica	Media	Error estandar	Intervalo	
			Lim inferior	Lim superior
1	5.93681192	0.14143913	5.90462032	5.96900351
2	6.00470596	0.13091737	5.97490912	6.0345028
3	6.20102717	0.13413848	6.17049721	6.23155713
4	5.86010602	0.13600787	5.82915058	5.89106146
5	6.27667326	0.12773727	6.24760021	6.30574631
6	6.14868327	0.13499559	6.11795823	6.17940831
7	6.02384314	0.14006417	5.99196449	6.05572179
8	5.93368243	0.13019772	5.90404938	5.96331548
9	5.81290795	0.12725495	5.78394468	5.84187122
10	5.99042988	0.1273659	5.96144136	6.01941841

- Teniendo en cuenta la distribución de los ingresos poblacionales, y basándose en este explique los resultados obtenidos en el punto cuatro. Justifique suficientemente su respuesta.

Como se aprecia en el gráfico 1. la distribución de los ingresos es casi uniforme, con algunas excepciones ya que existe un cambio brusco entre unos pocos individuos que devengan cero salarios mínimos con respecto a los que devengan un salario mínimo, y a su vez entre los que devengan once salarios mínimos con respecto a los que devengan más que esto. Esto implica que la estratificación que se plantea en punto 3 es ficticia, es decir no está acorde con los verdaderos ingresos poblacionales, ya que no existe heterogeneidad entre los estratos y es por esto que al obtener muestras estratificadas se incurre en un gran error de estimación, esto explica por qué las estimaciones del punto 4 b. son tan sesgadas, mientras que el muestreo aleatorio simple es el adecuado en este caso.

**Gráfico 1. Histograma de los ingresos poblacionales.**

